Assessment of Health State in Patients With Tinnitus: A Comparison of the EQ-5D and HUI Mark III

Iris H. L. Maes,[1,2,3] Manuela A. Joore,[1] Rilana F. F. Cima,[2,3] Johannes W. Vlaeyen,[3,4] and Lucien J. C. Anteunis[5]

Objectives: Expressing the outcomes of treatment in quality-adjusted life years is increasingly important as a tool to aid decision makers concerning the allocation of scarce resources within the health care sector. A quality-adjusted life year is a measure of life expectancy that is weighted by health-related quality of life. These weights are referred to as utility scores and are usually measured by multiattribute utility measures. Several studies found that different utility measures provide different estimates of the same person's level of utility. The aim of this study was to investigate which of two widely used utility measures, the EQ-5D and the HUI mark III, is preferred in a tinnitus population.

Methods: Baseline and follow-up data on EQ-5D and HUI mark III of 429 patients of a randomized controlled clinical trial, investigating cost-effectiveness of usual care versus specialized care of tinnitus, were included. Agreement, discriminative power, and responsiveness of the health state description and the utility scores were examined.

Results: Corresponding dimensions of the EQ-5D and HUI mark III showed large correlations; although ceiling effects were more frequently observed in the EQ-5D. Mean utility scores for EQ-5D (0.77; SD 0.22) and HUI mark III (0.64; SD 0.28) were significantly different (Wilcoxon signed ranks test, $p < 0.001$), and agreement was low to moderate (intraclass correlation coefficient = 0.53). Both health state description and utility scores of both measures discriminated between different severity groups. These groups were based on baseline scores of the Tinnitus Questionnaire. The HUI mark III had a higher ability than the EQ-5D to detect improved patients from randomly selected pairs of improved and unimproved patients.

Conclusion: This study shows that different utility measures lead to different health state descriptions and utility scores among tinnitus patients. However, both measures are capable of discriminating between clinically different groups. The HUI mark III is more responsive than the EQ-5D, and therefore preferred in a tinnitus population.

(Ear & Hearing 2011;32;428–435)

INTRODUCTION

Economic evaluation of health care is becoming increasingly important as a tool to aid decision makers concerning the allocation of scarce resources within the health care sector. To make a comparison between different conditions, there is a need to express the effectiveness of treatments in terms of quality-adjusted life years (QALYs). A QALY is a measure of life expectancy that is weighted by health-related quality of life. These weights are referred to as utility scores. Utilities are mostly determined indirectly, by using quality of life measures with pre-existing utility weights that can be attached to each permutation of responses. Unfortunately, the available multiattribute utility measures, such as the EuroQol-5D (EQ-5D) and the Health Utilities Index Mark III (HUI mark III), differ in the conceptualization and operationalization of health and the valuation method on which the utility scores are based. Several studies found that these two widely used multiattribute utility measures provide different estimates of the same person's level of utility (Brazier et al. 2004; Feeny et al. 2004; Hatoum et al. 2004; Marra et al. 2004; Espallargues et al. 2005; Marra et al. 2005; Grutters et al. 2007; Barton et al. 2008; Grieve et al. 2008).

As a result, these differences can lead to variations in the estimates of utilities and QALYs used in economic evaluations. The purpose of this study was to investigate systematic differences in health state descriptions and utility values obtained with the EQ-5D and the HUI mark III in tinnitus patients. Subjective tinnitus is the involuntary perception of the concept of a sound without the presence of an external source. It is a chronic condition that is highly prevalent, especially among hearing impaired individuals. Studies show a prevalence of 10 to 20% in the general population (Davis & El Refaie 2000). Among hearing impaired individuals, prevalence has been estimated at 75 to 80% (Adams et al. 1999). Among severe sufferers, tinnitus causes affective problems, major declines in concentration, sleeping difficulties, exhaustion, and problems in (re-)directing attention (Scott et al. 1990; Jastreboff et al. 1996; Erlandsson & Hallberg, 2000; Kroner-Herwig et al. 2003; El Refaie et al. 2004). These problems have detrimental effects on many areas of functioning, leading to a diminished quality of life.

The aim of this study was to determine which utility measure is preferred in a tinnitus population. To determine this, we will assess agreement and compare the discriminative power and responsiveness of both measures. Moreover, we will investigate the differences between the descriptive system and the utility scores of the EQ-5D and the HUI mark III. The article is structured as follows. First, we will introduce the dataset that was used. Next, we describe the EQ-5D and HUI mark III instruments and the analyses we performed. In the results section, we compare agreement, discriminative power, and responsiveness, for the health state description and the utility scores, of both instruments.

MATERIALS AND METHODS

Measures

The EQ-5D consists of a Visual Analog Scale and five questions, each representing a dimension of health-related quality of life: mobility, self-care, daily activities, pain/discomfort, and anxiety/depression (The EuroQol Group 1990). The Visual Analog Scale ranges from best (100) to worst (0) imaginable health

[1]Clinical and Medical Technology Assessment, Maastricht University Medical Centre, Maastricht, The Netherlands; [2]Adelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek, The Netherlands; [3]Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands; [4]Department of Psychology, University of Leuven, Belgium; and [5]Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands.

0196/0202/11/3204-0428/0 • Ear & Hearing • Copyright © 2011 by Lippincott Williams & Wilkins • Printed in the U.S.A.

state. On this scale, patients have to rate their own health state. The five questions have three response levels, which classify the severity of complaints on that specific dimension. The combination of levels over dimensions defines a universe of 243 unique health states (= 3 × 3 × 3 × 3 × 3). To determine an additive scoring function based on these different health states, Dolan (1997) derived preferences for these health states using a time trade-off task in a representative sample of the UK population of 2.997 respondents. In the time trade-off task, the respondents were asked to choose between two alternatives: remaining in a state of ill health for a period of time A or being restored to perfect health for a shorter period of time B. The utility of the ill health state equals B divided by A. The possible utility scores of the EQ-5D range from −0.59 to 1.0.

The HUI mark III consists of 17 questions that are used to compute eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain/complaints. Each question has five or six levels, and 972.000 possible health states can be formed by the levels of the different dimensions. A multiplicative utility scoring function was determined by deriving preferences from a Standard Gamble task and a Visual Analog Scale in a random sample of the Canadian general population of 504 respondents. In the Standard Gamble, respondents are asked to choose between two alternatives: remaining in a state of ill health for sure or an intervention with two possible outcomes, either restoring perfect health (p) or die immediately (1 − p). The risk a respondent is willing to take (p) represents their utility score. Possible utility scores range from −0.36 to 1.00 (Feeny et al. 2002) for the HUI mark III.

Differences between both measures may arise as a consequence of a difference in the conceptualization and the operationalization of health and a difference in the scoring algorithm to calculate utilities. The EQ-5D conceptualizes health as an overall construct containing physical, mental, and social functioning (Brooks 1996), whereas the HUI mark III focuses on health "within the skin," meaning that it purely focuses on the underlying level of impairment (Feeny et al. 1995). Also, the operationalization of health differs between the measures. The EQ-5D has a dimension focusing on the quality of the performance of daily activities, which the HUI mark III has not. The HUI mark III contains dimensions that are not present in the EQ-5D: vision, hearing, cognition, and dexterity. Especially the hearing and cognition dimensions are likely to be relevant in a population with tinnitus. Furthermore, the answering scales differ. The EQ-5D has three answering levels for each item, whereas the HUI mark III has five or six answering levels. Moreover, the answering scales are defined differently. The EQ-5D levels are defined as follows: no problems, moderate problems, or severe problems. The HUI mark III levels provide some explanation about what sort of complaints are associated with mild, moderate, or severe problems.

The scoring algorithm of both measures were derived using different methods. The EQ-5D UK tariff is based on time trade-off values, whereas the HUI mark III scoring algorithm is based on standard gamble and visual analog scale scores. It is described that the standard gamble leads to higher utility scores than the time trade-off, and the time trade-off leads to higher utility scores than the visual analog scale (Drummond et al. 2005). Also, the scoring algorithms were derived in different countries (UK versus Canada). Different cultures may value health in different ways (Knies et al. 2009).

The Tinnitus Questionnaire (TQ) (McCombe et al. 2001) is a measure of tinnitus-related distress and will serve as an external anchor to determine discriminative power and external responsiveness of the EQ-5D and HUI mark III. Based on the scores, the TQ patients can be classified into three different severity classes: mild tinnitus complaints, moderate tinnitus complaints, and severe tinnitus complaints.

Study Population and Data Collection

Data were collected as part of a ongoing randomized controlled trial investigating the effectiveness and cost-effectiveness of an integral multidisciplinary treatment for tinnitus versus care as usual (Cima et al. 2009). All patients referred to the Center of Audiology and Communication (Adelante, Hoensbroek) because of tinnitus complaints were included. Exclusion criteria were age younger than 18 yrs and not being able to read and write in Dutch.

The EQ-5D, HUI mark III, and the TQ were administered four times during a 12-mo period. At baseline (T0), all questionnaires were administered and completed during the first visit at the Audiological Centre. Respondents were assisted by a trained interviewer if requested. Three (T1), 8 (T2), and 12 mos (T3) after baseline, the patients were given access to an internet-based electronic environment to complete the battery of tests online. If patients were not able to complete the questionnaire online, a paper version was provided. This study includes only patients who had fully completed both questionnaires at baseline and after 3 mos. In the first 3 mos, the intervention focused on audiological rehabilitation (hearing aids and sound generators) and counseling.

Agreement

First, agreement in the health state descriptions obtained with the EQ-5D and HUI mark III were determined by computing frequency tables of the responses on the different dimensions of both measures. Kendall's tau was computed to determine the correlation between the corresponding dimensions of the measures: mobility (EQ-5D) and ambulation (HUI mark III), pain/complaints (EQ-5D) and pain (HUI mark III), and anxiety/depression (EQ-5D) and emotion (HUI mark III). Correlations were interpreted according to the following benchmarks: 0.1 to 0.3 was interpreted as small, 0.3 to 0.5 as medium, and >0.5 as large (Cohen 1988). Ceiling and floor effects of the EQ-5D and HUI mark III were studied by calculating the proportion of patients with either the best or worst health state description. For patients with a ceiling effect on the EQ-5D, the responses on the dimensions of the HUI mark III were presented in a frequency table.

To assess agreement between the health state utilities of both measures, a Wilcoxon signed rank test and a paired samples *t* test were used. Utility scores were tested for normality with a Kolmogorov-Smirnov test. The intraclass correlation coefficient (ICC) was computed based on a two-way mixed effect model, such that the subject effect was random and the instrument effect was fixed. Computations were performed at the individual patient level. An ICC <0.75 implies poor to moderate agreement and >0.75 implies good agreement (Gross Protney & Watkins 1993). Furthermore, Bland-Altman plots were computed with the difference between the utility scores from the EQ-5D and the HUI mark III on the Y axis and the mean of the utility scores on both measures on the X axis. All these analyses were performed on the baseline data.

Discriminative Power

Because there is also no gold standard for measuring health state utility, construct validity is usually determined by the discriminative power between clinically different groups. To determine the discriminative power of the health state descriptions of the EQ-5D and the HUI mark III, the frequencies of responses on the different dimensions were computed in clinically different groups based on TQ baseline scores. A score <30 on the TQ indicates mild tinnitus, a score between 31 and 46 indicates moderate tinnitus, and a score >47 indicates severe tinnitus. The discriminative power of the utility scores of the measures was determined by calculating the mean utility scores for these groups. Differences in mean scores were tested using analysis of variance or the Kruskal-Wallis test, depending on the distribution of the data.

Responsiveness

First, frequency tables of the health state descriptions of both measures were computed for patients who improved and patients who did not improve. Patients were classified as improved if the score on the TQ dropped 10 or more points from baseline to 3 mos.

Second, the health state utility scores were tested for internal and external responsiveness. Internal responsiveness is a distribution-based method that characterizes the ability of a measure to change over a particular prespecified time frame. External responsiveness is an anchor-based method that describes the relationship between change in a measurement and change in a reference measurement of health. The two most frequently used methods for determining internal responsiveness are the effect size (ES) and the standardized response mean (SRM) (Terwee et al. 2003). ES is calculated as the change in score divided by the SD of scores at baseline. The standardized response mean is the change in score divided by the SD of the change in score. Both the ES and the SRM were computed for those patients who improved and for patients who did not improve from baseline to 3 mos. They were interpreted using benchmarks for ES: 0.20 through 0.49 is interpreted as small, 0.50 through 0.79 as moderate, and ≥0.80 as large (Cohen 1988). External responsiveness was determined by a receiver operating curve (ROC) that determined how successfully a given score on the EQ-5D or the HUI mark III could discriminate patients who improved from those who did not improve. Improvement was defined as a decrease in the score on the TQ of 10 or more points and was the state variable. Test variables were the difference scores of the EQ-5D and the HUI mark III between baseline and the different follow-up measurements. The area under the ROC curve was interpreted as the probability of correctly identifying the improved patient from randomly selected pairs of improved and unimproved patients.

RESULTS

Study Population

Of the 429 patients included in the study, 428 completed the EQ-5D and the HUI mark III at baseline. Patients had a mean age of 54 yrs, and 62.2% were men. The mean hearing loss over the frequencies of 1000, 2000, and 4000 Hz was 31.50 (17.9). However, not all patients completed the follow-up measurement at 3 mos. The analyses of responsiveness in this article are based on the 319 patients who fully completed the EQ-5D and the HUI mark III at baseline and 3 mos follow-up. Patient characteristics are displayed in Table 1.

TABLE 1. Characteristics of the study population

Characteristics	Study population: Baseline	Study population: 3-mo follow-up
N	428	319
Age (yr)		
Mean (SD)	54.3 (11.8)	54.9 (11.4)
Range	20–85	20–85
Gender		
Male (%)	267 (62.2%)	203 (63.6%)
Living situation		
Living alone	85 (19.8%)	62 (19.4%)
Living together	343 (80.2%)	257 (80.6%)
Mean PTT at 1, 2, and 4 kHz (in dB)		
Right	30.0	30.8
Left	31.0	32.4
Overall	31.5	31.8
Duration of tinnitus complaints		
<1 yr (%)	126 (29.4%)	92 (28.8%)
1–5 yrs (%)	167 (39.0%)	122 (38.2%)
5–10 yrs (%)	60 (14.0%)	46 (14.5%)
>10 yrs (%)	75 (17.6%)	59 (18.5%)

PTT, pure-tone threshold.

Agreement

The frequency tables in Figure 1 show that most patients had complaints on the dimensions pain/complaints, anxiety/depression, and daily activities of the EQ-5D. For the HUI mark III, most patients had complaints in the pain, emotion, cognition, and hearing dimension (Fig. 1). With regard to vision, the HUI mark III shows that approximately 80% of the patients wear glasses. There were large positive correlations (Kendall's tau > 0.50) between mobility (EQ-5D) and ambulation (HUI mark III) (Kendall's tau = 0.595) and anxiety/depression (EQ-5D) and emotion (HUI mark III) (Kendall's tau = 0.527). There was a moderate correlation between pain/complaints (EQ-5D) and pain (HUI mark III) (Kendall's tau = 0.487). In both measures, floor effects were not observed. Ceiling effects were more frequent in the EQ-5D health state descriptions. The frequency table of Figure 2 shows that patients with a ceiling effect on the EQ-5D report most problems on dimensions that are only present in the HUI mark III, such as hearing and cognition. However, they also show some problems on HUI mark III dimensions that correspond with EQ-5D dimensions, such as emotion and pain.

Both the EQ-5D and the HUI mark III utility scores were not normally distributed ($p < 0.001$). The mean utility score for HUI mark III was 0.64 and the mean utility score for the EQ-5D was 0.77. The utility scores were significantly differently distributed (Wilcoxon signed ranks test, $p < 0.001$). The ICC for agreement was 0.53 (95% CI 0.45–

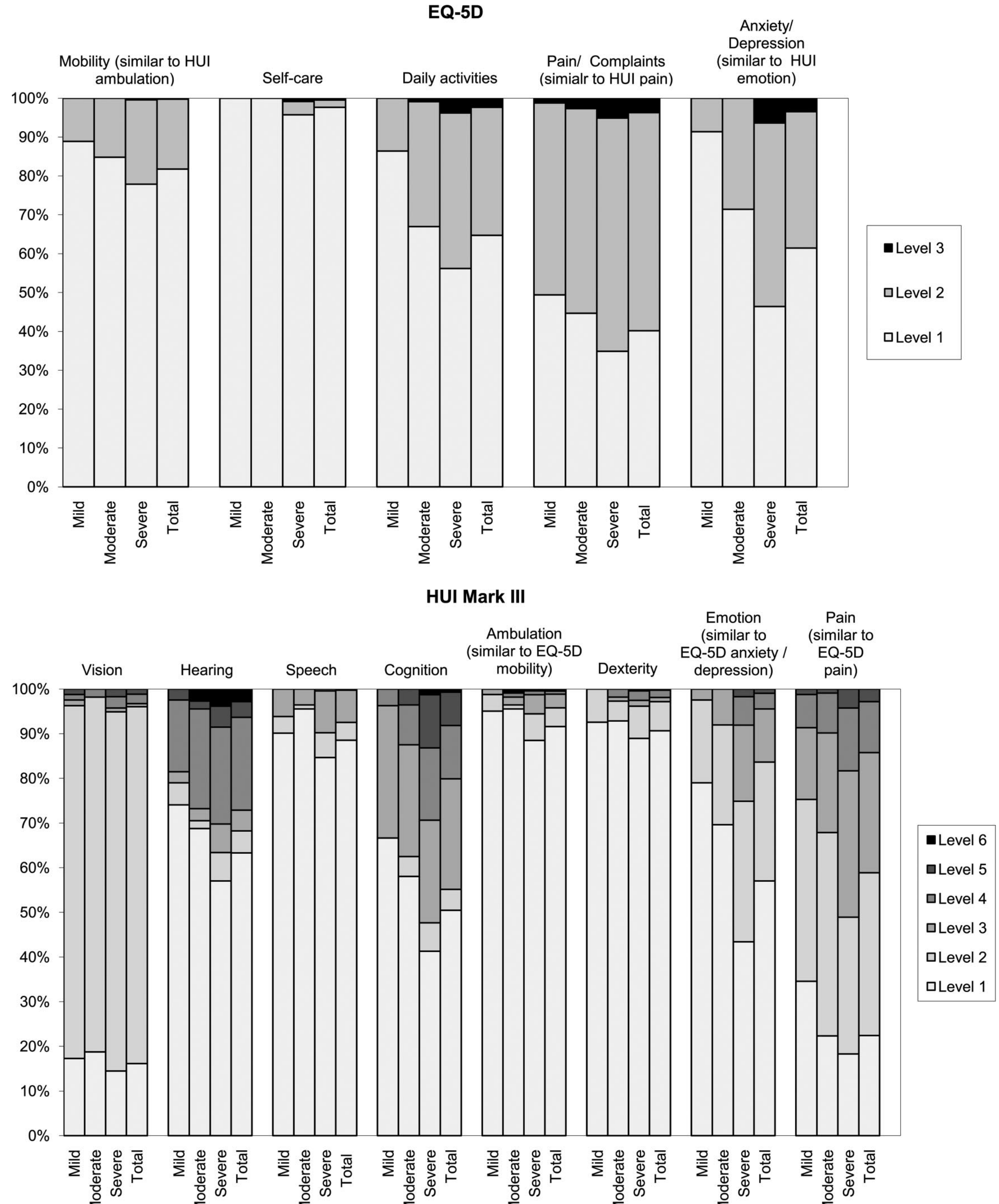

Fig. 1. Percentage of responses on dimensions of EQ-5D and HUI mark III at baseline for mild (N = 81), moderate (N = 112), and severe (N = 235) tinnitus and the total group (N = 428).

0.60), which implies a poor to moderate agreement. Bland-Altman plots (Fig. 3) show that the difference between the EQ-5D and HUI mark III utility scores are larger when the mean utility scores of both measures increases. Furthermore, it shows that the HUI mark III overall has lower utility scores than the EQ-5D.

Construct Validity

Patients were divided into three clinically different groups based on the TQ scores. These showed that 81 patients (19%) had mild tinnitus complaints, 112 (26%) patients had moderate tinnitus complaints, and 235 (55%) patients had severe tinnitus complaints at baseline. A one-way analysis of variance showed that there were no differences between these groups with regard to age, sex, or hearing loss.

With regard to the EQ-5D, the daily activities and the anxiety/depression dimensions distinguish best between mild, moderate, and severe groups, with more complaints in groups with more severe tinnitus (see Fig. 1). The mobility and pain/complaints dimension also distinguish between the differ-

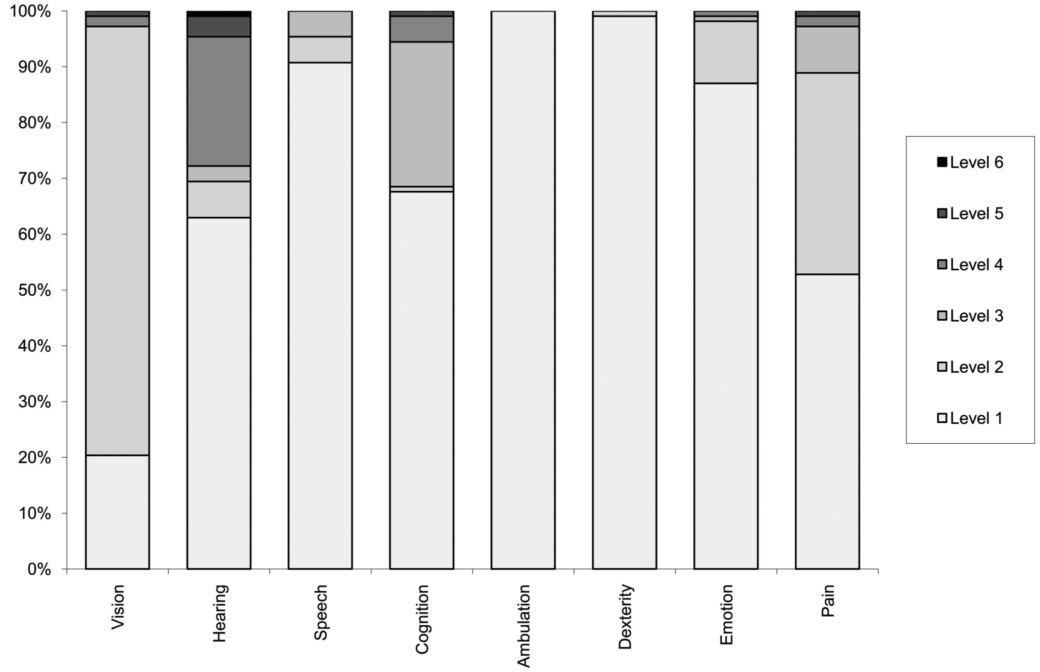

Fig. 2. Ceiling effects of EQ-5D. Distribution of responses (%) on the HUI mark III dimensions for those with EQ-5D = 11111 (N = 108).

ent severity groups. The EQ-5D self-care dimension shows no complaints in the mild and moderate groups and some complaints in the severe group. With regard to the HUI mark III, the dimensions hearing, cognition, emotion, and pain distinguish between groups with mild, moderate, and severe tinnitus. In the other dimensions, patients report little or no complaints.

Table 2 provides an overview of the mean utility scores for the different clinical groups as measured by the EQ-5D and the HUI mark III. Both utility measures discriminate between clinically different groups (analysis of variance, $p < 0.001$; Kruskal-Wallis test, $p < 0.001$). Groups with more severe tinnitus had significantly lower mean utility scores at both baseline and after 3 mos.

Responsiveness

From baseline to 3 mos, 112 patients improved at least 10 points on the TQ, and 207 patients showed no improvement.

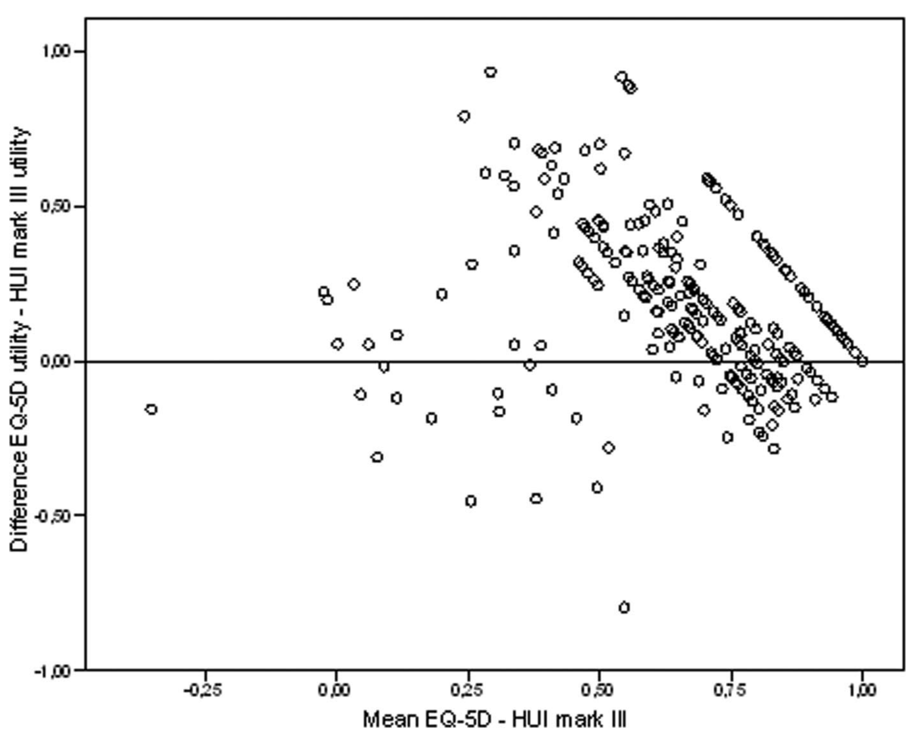

Fig. 3. Bland-Altman plot of EQ-5D utility score versus HUI mark III utility score (N = 428).

TABLE 2. Mean utility scores for EQ-5D and HUI mark III

		Baseline				3-mo follow-up			
	TQ	N	% Ceiling effect	Mean* (SD)	Median†	N	% Ceiling effect	Mean* (SD)	Median†
EQ-5D	Mild	81	43.2	0.87 (0.15)	0.79	55	47.3	0.86 (0.16)	0.81
	Moderate	112	33.9	0.82 (0.17)	0.80	86	38.4	0.84 (0.18)	0.80
	Severe	235	14.9	0.71 (0.24)	0.76	178	16.3	0.71 (0.25)	0.73
	Total	428	25.2	0.77 (0.22)	0.80	319	27.6	0.77 (0.23)	0.80
HUI mark III	Mild	81	6.2	0.79 (0.18)	0.85	55	1.8	0.88 (0.18)	0.84
	Moderate	112	2.7	0.82 (0.17)	0.78	86	2.3	0.73 (0.24)	0.79
	Severe	235	0.4	0.55 (0.30)	0.55	178	1.1	0.54 (0.30)	0.58
	Total	428	2.1	0.64 (0.28)	0.70	319	1.6	0.63 (0.28)	0.70

* *All statistically significant, ANOVA* ($p < 0.001$).
† *All statistically significant, Kruskal-Wallis test* ($p < 0.001$).

Figure 4 shows that after 3 mos, there are slightly less complaints on each of the dimensions of the EQ-5D. In the HUI mark III dimensions, patients have fewer complaints on the pain, emotion, and cognition dimension. With regard to the hearing dimension, patients show more complaints. There were no changes in complaints on the dimensions of the EQ-5D in patients who did not show improvement in the first 3 mos. In the HUI mark III, this group showed slightly more complaints on the hearing, emotion, and cognition dimension.

The mean change in utility scores in patients who did not improve on the TQ was not statistically significant for the EQ-5D or the HUI mark III. The mean change in utility scores for improved patients from baseline to 3 mos was statistically significant for the both HUI mark III (Wilcoxon signed ranks test, $p < 0.01$) and the EQ-5D (Wilcoxon signed ranks test, $p < 0.05$). The ESs of the change in health state utilities from baseline to 3 mos were 0.19 (SD 0.84) for the EQ-5D and 0.21 (SD 0.75) for the HUI mark III. The SRM of the change in score from baseline to 3 mos was 0.22 (SD 0.94) for the EQ-5D and 0.25 (SD 0.97) for the HUI mark III. Although all these ESs are small, the HUI mark is slightly more responsive than the EQ-5D in patients with tinnitus. With regard to the external responsiveness, the area under the curve was 0.61 for the HUI mark III and 0.58 for the EQ-5D (Fig. 5).

DISCUSSION

We investigated agreement, construct validity, and internal and external responsiveness of the health state description and utility scores based on the EQ-5D and the HUI mark III in patients with tinnitus. The results of this study provide insight into the differences between the two widely used utility measures in patients with tinnitus. Our main findings are the following.

First, although corresponding dimensions showed large correlations, in the EQ-5D, health state description ceiling effects were much more frequently observed. Ceiling effects of the EQ-5D are already reported in other studies (Bharmal & Thomas 2006; Grutters et al. 2007; Luo et al. 2009). Because of the presence of these ceiling effects, it was suggested not to use the EQ-5D in relatively healthy populations (Kopec &

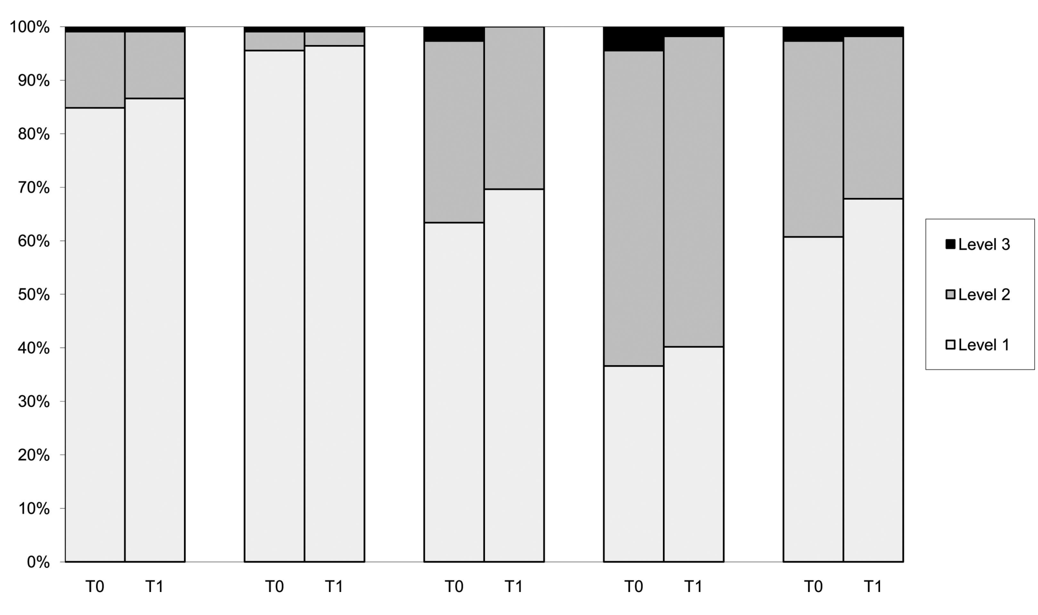

Fig. 4. Frequency of dimensions of EQ-5D and HUI mark III of patients who improved from T0 to T1 (N = 112).

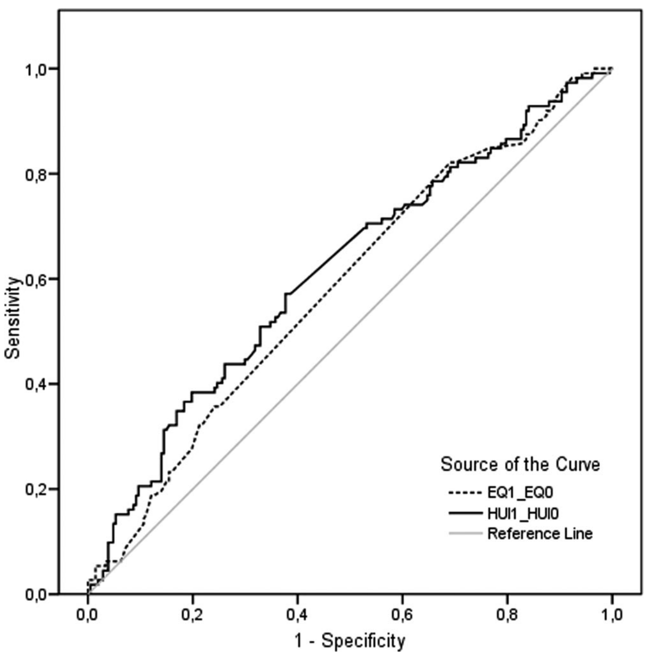

Fig. 5. External responsiveness of EQ-5D and HUI mark III from baseline to 3 mos (N = 319).

Willison 2003). Patients with perfect health according to the EQ-5D reported substantial problems on both corresponding and noncorresponding dimensions of the HUI mark III. From this we can conclude that both the operationalization of health in items, as the answering scale used, seem to cause differences between these two measurements. It also illustrates one of the disadvantages of the EQ-5D for otherwise healthy populations but characterized (in part) by a cognitive and/or sensory dysfunction.

Second, the utility scores of the EQ-5D were higher than the utility scores of the HUI mark III. These findings are in line with the literature (Oostenbrink et al. 2002; Marra et al. 2005; Grutters et al. 2007; Luo et al. 2009). Part of the differences between utility scores can be explained by the differences in health state descriptions, but it is unclear to what amount. Differences could also result from the differences in the utility scoring function. Although in general the time trade-off (used for the EQ-5D) leads to lower scores than the standard gamble (used for the HUI mark III), the HUI mark III utility scores were considerably lower than the EQ-5D utility scores in this study. The utility scoring function of the EQ-5D is additive, assuming no interaction for preferences among attributes at all. The HUI mark III uses a multiplicative scoring function, with the effect that the level of impairment in one dimension is dependent on the level of impairment in another dimension. The utility scores of the HUI mark III are expected to be lower, because they take into account the effect of comorbidity.

Third, both measures have discriminative power regarding the health state description. The corresponding dimensions show the most differences between clinically different groups based on the severity of the tinnitus. The HUI mark III also shows a large effect on the hearing and cognition dimension. This is in line with our expectation, because tinnitus is more prevalent in patients with hearing complaints. Furthermore, tinnitus leads to problems in concentration and redirecting attention.

Fourth, health state utility scores of both measures also have discriminative power in groups with more severe tinnitus. These findings are in line with other studies that reported discriminative power of both measures (Marra et al. 2004; Fisk et al. 2005). One study reported discriminative power of only the HUI mark III (Grutters et al. 2007). In this study, the utility scores of the HUI mark III are lower than the EQ-5D utility scores for each tinnitus severity group.

Fifth, both the EQ-5D and the HUI mark III show some improvement in the different dimensions from baseline to 3 mos. The HUI mark III shows more complaints in the hearing dimension. This could be a result of the fact that treatment in the first 3 mos is especially aimed at audiological rehabilitation, drawing their attention to the hearing loss and complaints.

Sixth, the HUI mark III and the EQ-5D both measured statistically significant change in the mean utility scores of improved patients after the first 3 mos, in which the first level of treatment was finished. In two other studies that compared both measures in a population of patients with hearing loss, only the HUI mark III was able to measure statistically significant improvement (Barton et al. 2004; Grutters et al. 2007). An integral multidisciplinary treatment for tinnitus is aimed at improving quality of life, while hearing aid fitting will be most noticeable in the hearing dimension that is only represented in the HUI mark III. In this study, the HUI mark III had a higher ability to detect improved patients, from randomly selected pairs of improved and unimproved patients. A possible explanation for the lack of sensitivity to measure change in the EQ-5D is the occurrence of a ceiling effect in 25% of the population at baseline. If such a large proportion of patients report perfect health at baseline, it is unlikely to find a considerable utility gain from any intervention.

A shortcoming of this study is the fact that it did not allow us to estimate an instrument order effect. In all subjects, the TQ was administered first, followed by the HUI mark III and the EQ-5D. The extent to which a change in instrument order would influence differences in the health state description and health state utility is unknown.

In conclusion, the results on agreement, construct validity, and responsiveness show substantial differences between the health descriptions, as well as the utility scores, between the EQ-5D and the HUI mark III. Differences in conceptualization and operationalization of health explain part of the differences in the utility scores between both measures. It remains unclear how the scoring differences of both measures are responsible for the differences between the utility scores. According to the results of this study, both the EQ-5D and the HUI mark III can be used in a tinnitus population, although researchers should be aware of the possible ceiling effects of the EQ-5D. This contributes to the evidence that there may not be a superior instrument for measuring health state utility. For now, researchers should use a measurement tool that best fits the condition under investigation. Despite considerable overlap between both measures, we recommend the HUI mark III as the tool of preference in patients with tinnitus, because it is the most sensitive to change in the condition and is less affected by ceiling effects.

ACKNOWLEDGMENTS

Address for correspondence: Iris H. L. Maes, Maastricht University Medical Centre, PO Box 5800, 6202 AZ Maastricht, The Netherlands. E-mail: iris.maes@mumc.nl.

Received December 1, 2009; accepted September 10, 2010.

REFERENCES

Adams, P. F., Hendershot, G. E., Marano, M. A. (1999). Current estimates from the National Health Interview Survey, 1996. *Vital Health Stat*, *10*, 1–203.

Barton, G. R., Bankart, J., Davis, A. C., et al. (2004). Comparing utility scores before and after hearing-aid provision: Results according to the EQ-5D, HUI3 and SF-6D. *Appl Health Econ Health Policy*, *3*, 103–105.

Barton, G. R., Sach, T. H., Avery, A. J., et al. (2008). A comparison of the performance of the EQ-5D and SF-6D for individuals aged >or= 45 years. *Health Econ*, *17*, 815–832.

Bharmal, M., & Thomas, J., III. (2006). Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health*, *9*, 262–271.

Brazier, J., Roberts, J., Tsuchiya, A. et al. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*, *13*, 873–884.

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*, 53–72.

Cima, R., Joore, M., Maes, I., et al. (2009). Cost-effectiveness of multidisciplinary management of Tinnitus at a specialized Tinnitus centre. *BMC Health Serv Res*, *9*, 29.

Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, A., & El Refaie, A. (2000). *Epidemiology of Tinnitus.* San Diego, CA: Singular.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Med Care*, *35*, 1095–1108.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., et al. (2005). *Methods for the Economic Evaluation of Health Care Programmes.* New York, NY: Oxford University Press.

El Refaie, A., Davis, A., Kayan, A., et al. (2004). A questionnaire study of the quality of life and quality of family life of individuals complaining of tinnitus pre- and post-attendance at a tinnitus clinic. *Int J Audiol*, *43*, 410–416.

Erlandsson, S. I., & Hallberg, L. R. (2000). Prediction of quality of life in patients with tinnitus. *Br J Audiol*, *34*, 11–20.

Espallargues, M., Czoski-Murray, C. J., Bansback, N. J., et al. (2005). The impact of age-related macular degeneration on health status utility values. *Invest Ophthalmol Vis Sci*, *46*, 4016–4023.

Feeny, D., Furlong, W., Boyle, M., et al. (1995). Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics*, *7*, 490–502.

Feeny, D., Furlong, W., Torrance, G. W., et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care*, *40*, 113–128.

Feeny, D., Wu, L., Eng, K. (2004). Comparing short form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: Results from total hip arthroplasty patients. *Qual Life Res*, *13*, 1659–1670.

Fisk, J. D., Brown, M. G., Sketris, I. S., et al. (2005). A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *J Neurol Neurosurg Psychiatry*, *76*, 58–63.

Grieve, R., Grishchenko, M., Cairns, J. (2009). SF-6D versus EQ-5D: Reasons for differences in utility scores and impact on reported cost-utility. *Eur J Health Econ*, *10*, 15–23.

Gross Protney, L., & Watkins, M. P. (1993). *Foundations of Clinical Research.* Norwalk, CT: Appleton & Lange.

Grutters, J. P., Joore, M. A., van der Horst, F., et al. (2007). Choosing between measures: Comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. *Qual Life Res*, *16*, 1439–1449.

Hatoum, H. T., Brazier, J. E., Akhras, K. S., et al. (2004). Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. *Value Health*, *7*, 602–609.

Jastreboff, P. J., Gray, W. C., Gold, S. L. (1996). Neurophysiological approach to tinnitus patients. *Am J Otol*, *17*, 236–240.

Knies, S., Ament, A., Evers, S., et al. (2009). The transferability of economic evaluations: Testing the model of Welte. *Value Health*, *12*, 730–738.

Kopec, J. A., & Willison, K. D. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol*, *56*, 317–325.

Kroner-Herwig, B., Frenzel, A., Fritsche, G., et al. (2003). The management of chronic tinnitus: Comparison of an outpatient cognitive-behavioral group training to minimal-contact interventions. *J Psychosom Res*, *54*, 381–389.

Luo, N., Johnson, J. A., Shaw, J. W., et al. (2009). Relative efficiency of the EQ-5D, HUI2, and HUI3 index scores in measuring health burden of chronic medical conditions in a population health survey in the United States. *Med Care*, *47*, 53–60.

Marra, C. A., Esdaile, J. M., Guh, D., et al. (2004). A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care*, *42*, 1125–1131.

Marra, C. A., Woolcott, J. C., Kopec, J. A., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med*, *60*, 1571–1582.

McCombe, A., Baguley, D., Coles, R., et al. (2001). Guidelines for the grading of tinnitus severity: The results of a working group commissioned by the British Association of Otolaryngologists, Head and Neck Surgeons, 1999. *Clin Otolaryngol Allied Sci*, *26*, 388–393.

Oostenbrink, R., Moll, H. A., Essink-Bot, M. L. (2002). The EQ-5D and the Health Utilities Index for permanent sequelae after meningitis: A head-to-head comparison. *J Clin Epidemiol*, *55*, 791–799.

Scott, B., Lindberg, P., Melin, L., et al. (1990). Predictors of tinnitus discomfort, adaptation and subjective loudness. *Br J Audiol*, *24*, 51–62.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., et al. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res*, *12*, 349–362.

The EuroQol Group. (1990). EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy*, *16*, 199–208.